



Estimating the Acute Gastrointestinal Disease Burden Attributable to Microbes in Drinking Water

Michael Messner¹, Susan Shaw¹, Valerie Blank¹, Stig Regli¹, and Jeff Soller²

¹USEPA Office of Ground Water and Drinking Water, Standards and Risk Management Division

²Risk Policy Fellow, American Association for the Advancement of Science



BACKGROUND

A 12-month blinded household intervention in Davenport, IA, to determine the fraction of GI illness attributable to drinking water, is now published. The study was an example of collaboration between EPA's research labs, its Office of Ground Water and Drinking Water (OGWDW), and CDC's National Center for Infectious Disease (NCID).

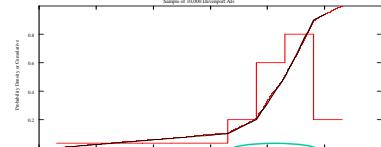
This poster presents an approach to estimating, on a national level, the incidence of GI illness cases that may be attributed to drinking water. It assumes that the incidence of drinking water-related GI illnesses per person-year, e.g., the attributable incidence (AI), is lognormally distributed across communities and considers Davenport's placement within the national AI distribution.

PRIOR INFORMATION

Davenport's AI

The figure below describes the Team's prior beliefs about Davenport's incidence of GI illness attributable to drinking water. We believed there was a high likelihood that a statistically significant effect would be found by the household intervention study in Davenport. This belief was supported by 1) the results of previous household intervention studies in Laval, Canada (Payment et al., 1991, 1997), 2) our understanding of the relative levels of microbial risk based on comparisons of published information on source water quality, water treatment and distribution system microbial risk of the Laval system during the two study periods, and 3) what we knew about Davenport's source water from published studies, and its level of treatment (information on treatment type and compliance with federal microbial regulations).

In the figure, the red step function is the prior density and the black curve is the resulting (cumulative) distribution function. Beneath the black curve is the distribution of a sample (size 10,000) drawn from this prior. It coincides very nicely, as one would expect with such a large sample.



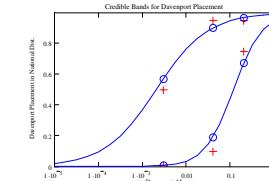
Prior to viewing any data, the Team expected Davenport's AI to fall somewhere in this range with probability 0.8.

Davenport's Placement

The team was able to express its belief about Davenport's placement (e.g. percentile) in the US distribution, conditioned on knowing the magnitude of Davenport's AI.

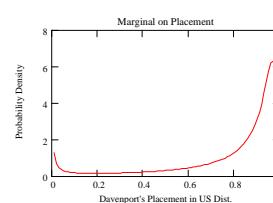
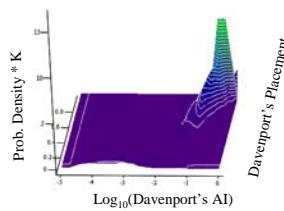
- If we were to know that Davenport's AI was truly 0.2 cases per person-year, then we would be highly confident (95% level) that Davenport's placement would be in the upper half of the US distribution, specifically, in the interval (0.75, 0.95).
- If we knew that Davenport's AI was truly 0.04 cases per person-year, then we'd expect its placement to fall in the interval (0.1, 0.95).
- If we knew that Davenport's AI was truly 0.003 cases per person-year, then we'd expect its placement to fall in (0.01, 0.5).
- If we knew that Davenport's AI was truly 1 case per person-year, then we'd expect its placement to fall above 0.95.

These prior beliefs were modeled as follows: Following logit transform ($\text{logit}(p) = \ln(\text{odds ratio})$), the Team's upper credible limits appeared to vary as a linear function of $\ln(\text{AI})$. The lower limits also appeared to vary as another linear function of $\ln(\text{AI})$. The possibility of normal uncertainty was tested by showing the group the 95% credible intervals that result from applying the least-squares estimated lines to values of AI between and beyond those specified by the group. Additionally, the group examined the values estimated for calibration points 1 through 4, above. All appeared to be satisfactory, so the least-squares fitted lines were used for interpolation and extrapolation over the range of interest (Davenport's AI in the range of 0.00002 to 2 cases per person-year). The figure to the right describes the upper and lower credible bands over that range (blue curves). The blue circles are estimates for the selected values of AI and the red '+'s are the team's expressed values for those same AIs.



Joint Distribution (Davenport's AI, Davenport's Placement)

The resulting joint distribution of AI and placement is shown in the 3-D plot below. Integrating across AI, a marginal for placement was found to be that described in the figure below and to the right. Notice the bulk of the prior probability mass is in the upper range. This is consistent with the Team's expressed belief that Davenport's AI was sufficiently great to produce a statistically significant finding by the household intervention study. The peak near zero reflects the team's belief in the possibility that Davenport's AI is small or very small and therefore near the lower end of the US distribution.



Variability of AI in the US

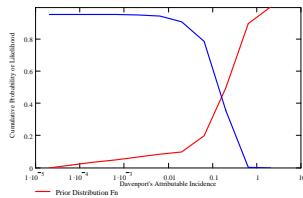
Independent of Davenport's placement or AI, the team believed that the central 95% of US AIs could span a range of 100 to 100,000. On a log scale ($\log(100) = 2$ to $\log(100,000) = 5$), we felt any equal-sized intervals were equally likely to contain the true log(range). The US distribution of AI was therefore modeled as lognormal with central 95% interval width uniform on a log scale (between 2 and 5).

DISCLAIMER: The findings and conclusions in this poster are those of the authors and do not necessarily represent the views of the Environmental Protection Agency.

NEW INFORMATION

Davenport household intervention Study

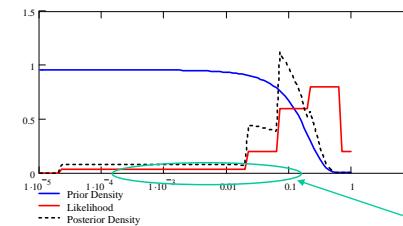
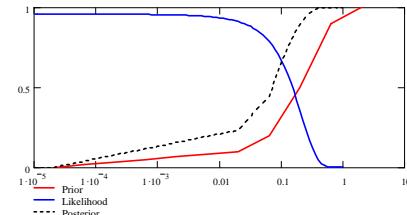
Based on estimated highly-credible acute gastrointestinal illness rates in the group receiving treated and untreated tap water, the likelihood function for the AI rate is normal with mean = 0.080 and standard deviation 0.191. The negative mean is due to finding a higher illness rate in the group receiving tap water. The figure below displays this new likelihood function together with the Team's prior distribution function. The study evidence is strong where our prior is weak, e.g., AI is less than 0.01 cases per person-year. While our prior rejects the notion that AI could be less than 10.5, the new information is strong evidence that AI does not exceed 0.5 cases per person-year.



COMBINING PRIOR AND NEW INFORMATION

Bayesian Updating

The posterior density for Davenport's AI is proportional to the product of the Prior Density and Likelihood. In the first figure below, the posterior distribution function (dashed line) is similar to the prior, but shifted to the left, based on the new information. The second figure below displays the prior and posterior density functions. The Team has taken to calling the posterior an "Opera House Density."



After processing the data, the Team's 80% credible interval has shifted to the left (compared to the Team's prior).

Estimating the National Average AI = NE (National Estimate)

Sampling Davenport's AI from the Bayesian posterior, sampling Davenport's placement from the conditional distribution (placement, given AI), and independently sampling 10,000 values for log-dispersion of the national distribution, completes the uncertain information as 10,000 ordered triplets (Davenport AI, Davenport Placement, Log Dispersion). For each of these 10,000 "realizations," we compute the national estimate as the expected attributable incidence rate:

$$\sigma_{\text{sim}} := \sqrt{\ln(10) \text{LOGS}_{\text{sim}}}$$

$$\mu_{\text{sim}} := \ln(10) Y_{\text{sim},0} - \text{qnorm}(Y_{\text{sim},1}, 0, \sigma_{\text{sim}})$$

$$\text{NE}_{\text{sim}} := \frac{\int_{-\infty}^{\text{Limit}} \text{norm}(x, \mu_{\text{sim}}, \sigma_{\text{sim}})^2 dx}{\text{pnorm}(\text{Limit}, \mu_{\text{sim}}, \sigma_{\text{sim}})}$$

Summary Results

The end result of the above is a set of 10,000 "National Estimates," each the expected Attributable Incidence, given a specific Davenport AI, Davenport Placement, and Log Dispersion of AI across the US. These 10,000 vary because of uncertainty in these three parameters. Summary statistics are as follows:

$$\text{mean}(\text{NEI}) = 0.119$$

$$\text{sortNEI}_{0.025-\text{SIMS}-1} = 0.013$$

$$\text{sortNEI}_{0.975-\text{SIMS}-1} = 0.377$$

The National Estimate is believed to fall between 0.01 and 0.38 cases per person-year with probability 0.95. A mean estimate is 0.12 cases per person-year, which would be approximately 15% of the nation's total acute GI disease burden.

88% of the uncertainty variance in the National Estimate can be explained by location parameter μ , which can be regarded as the median AI rate, which is Based on Davenport's placement and AI. This suggests that a study to better define the median AI rate would be more valuable than a study to better define the degree of variability in AI rates across the US.

REFERENCES

- Cofield, J., T.J. Wade, et al. (in press). A Randomized, Controlled Trial of In-Home Drinking Water Intervention to Reduce Gastrointestinal Illness. *Am. J. Epidemiol.* 2005; 161: 472-482.
- Payment, P., L. Richardson, et al. (1991). A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *Am. Jour. Public Health* 81(6):703-708.
- Payment, P., J. Siemiatyczik, et al. (1997). A prospective epidemiological study of gastrointestinal health effects due to the consumption of drinking water. *International Journal of Environmental Health Research* 7:5-31.
- Imhoff, B., D. Mors, et al. (2004). Burden of self-reported acute diarrheal illness in FoodNet surveillance areas, 1998-1999. *Clinical Infectious Diseases* 38(Suppl 3):S219-S226.



epascienceforum
Collaborative Science
for Environmental Solutions



2005
epascience